

# cigref

LES PROBLÈMES POSÉS  
PAR LA GESTION DE GRANDS FICHIERS

RAPPORT DE SYNTHÈSE  
GROUPE "LOGICIELS"

Février 1979

Réf. 79 149

**Février 1979**

**LES PROBLEMES POSES  
PAR LA GESTION DES GRANDS FICHIERS**

**- RAPPORT DE SYNTHÈSE DU GROUPE "LOGICIELS ET MÉTHODOLOGIE" -**

## SOMMAIRE

	Page
INTRODUCTION.....	2
CONDITIONS GÉNÉRALES .....	3
La notion de grand fichier .....	3
Nature et enveloppe des problèmes posés .....	5
ANALYSE DES SYSTÈMES DÉVELOPPÉS .....	8
(dans les entreprises du CIGREF)	
Description sommaire des systèmes présentés .....	8
Caractéristiques de l'organisation des données .....	9
VIE OPÉRATIONNELLE DU GRAND FICHER .....	13
Gestion des données .....	13
Administration de la Base .....	14
<b>Conclusion</b> .....	<b>15</b>
CONCLUSIONS GÉNÉRALES .....	16

## INTRODUCTION

La gestion des grands fichiers informatiques pose des problèmes et le thème est revenu à plusieurs reprises dans les réponses à l'enquête organisée par le CIGREF, auprès des responsables informatiques, début 1977.

Le rapport de synthèse proposé ici repose sur l'examen des difficultés rencontrées lors de la mise en œuvre et de l'exploitation de neuf grands fichiers dans sept entreprises membres du CIGREF.

Il importe d'abord de dire que le facteur "taille" n'est pas le seul significatif, bien qu'étant le révélateur de problèmes qu'il amplifie.

En effet, si toutes les applications décrites mettaient en œuvre des fichiers de plus de 200 mégaoctets, il est apparu que c'était essentiellement les fichiers de base des entreprises, leurs fichiers opérationnels, qui posaient des problèmes aux informaticiens attachés à répondre efficacement aux soucis quotidiens des utilisateurs.

Ces soucis étant forcément liés à la spécificité des activités des entreprises, il est aussi naturel de constater des différences notables entre les solutions apportées, qui sont d'ailleurs évolutives, les besoins se précisant à l'usage.

Notamment les questions d'accès tant sous les aspects physiques que logiques, ont posé à tous des problèmes d'intégrité, de cohérence, de confidentialité de l'information, donc des problèmes de structuration des données; question essentielle aussi au niveau des coûts informatiques et délais de réponse.

C'est donc le vaste problème de l'informatique opérationnelle dans l'entreprise, de son intégration dans sa vie quotidienne, à travers des applications fondamentales, qui était posé.

Est-il nécessaire, dans ces conditions, de préciser que ce rapport ne prétend pas résoudre les problèmes posés. Les utilisateurs et concepteurs de ces grands fichiers en sont conscients lorsqu'ils préparent une nouvelle version de leurs systèmes après en avoir observé le fonctionnement.

Cependant, un examen attentif des orientations prises dans des contextes voisins permettra, à défaut de panacée, d'éviter des voies peut-être prometteuses, qui se révèlent être sans lendemain dans les conditions actuelles de la technique informatique.

## CONSIDÉRATIONS GÉNÉRALES

### LA NOTION DE GRAND FICHIER

A la suite d'un questionnaire adressé avant l'été 1977 aux responsables informatiques des sociétés membres du CIGREF, il est apparu que les problèmes posés par l'organisation et la gestion des grands fichiers figuraient parmi les cinq ou six sujets d'intérêt majeur des responsables dans de nombreuses entreprises. L'étude et la discussion de ces problèmes a été confiée au groupe "Logiciels et Méthodologie".

Il importe de pouvoir d'abord définir ce qu'on entend par *grand fichier*, cette qualification ayant une valeur essentiellement relative.

En pratique, un fichier est jugé *grand* quand il pose des problèmes de maniabilité eu égard aux difficultés de sa mise en œuvre, de sa gestion et de son exploitation par rapport aux moyens que l'entreprise peut aligner. Soulignons tout de suite que les moyens des entreprises du CIGREF ne sont pas précisément des plus minces.

On retiendra donc les différents critères d'appréciation suivants à pondérer les uns par rapport aux autres :

#### - Volume des données

Interviennent à la fois le volume total et la dimension de chacun des articles qui influent tous deux directement sur les difficultés d'exploitation.

#### - Structure des données

Plus peut-être que le volume des données, intervient la complexité de leur structuration logique.

#### - Nature des mouvements subis par le fichier en exploitation

Les temps d'exploitation dépendent directement à la fois de la fréquence des mouvements et de leur diversité ; diversité qui peut amener à dupliquer certaines données même à l'intérieur d'une base, voire à éclater la base.

#### - Diversité des utilisateurs du fichier

Le plus souvent, un grand fichier sert de support à des applications différentes dans les entreprises. La multiplicité des accès qui en découle pose des problèmes complexes : qui accède ? à quelles informations pour quoi faire (consulter, mettre à jour) ?

- Mode d'exploitation du fichier

En général, un grand fichier est *i* la fois exploité en traitement par lot et en temps réel.

Mais le problème délicat est celui de la façon dont il est mis *i* jour

- Directement en temps réel par un ou plusieurs utilisateurs,
- en temps réel apparent (fichier intermédiaire d'attente de contrôle et validation),
- en traitement différé.

Il importe aussi de distinguer dès la conception les informations faisant l'objet de consultations fréquentes et celles souvent mises *i* jour.

- Mode de gestion et réorganisation du grand fichier

Dans la mesure où il est utilisé conjointement par plusieurs utilisateurs, le grand fichier pose des problèmes de gestion et de réorganisation, qui ne se posent pas dans le cas de fichiers mono-utilisateur.

Cette multiplicité des utilisateurs amène souvent *i* distinguer

Une fonction *gestion du contenant* (intégrité et sécurité d'accès par exemple) qu'on appellera *administration*,

- une fonction *gestion du contenu* qu'on appellera *gestion*.

Les systèmes de gestion de bases de données (SGBD) ont des spécifications qui permettent de répondre au problème posé, mais l'expérience montre qu'ils sont lourds à mettre en œuvre et coûteux en exploitation.

Il est donc nécessaire, dans certains cas, de développer des architectures de données, et des méthodes d'accès spécifiques aux applications envisagées, en ne retenant, parmi les possibilités des SGBD, que les fonctionnalités nécessaires.

Nous n'avons pas abordé les problèmes de recherches documentaires qui sont *i* la fois spécifiques et assez bien résolus.

On voit donc que la notion de grand fichier ne doit pas être restreinte aux seules bases de données. Réciproquement toutes les bases ne sont pas nécessairement des grands fichiers. Il est néanmoins certain que la notion de grand fichier amène *i* poser le problème spécifique des bases de données.

C'est la synthèse des expériences de sept sociétés du CIGREF qu'il est proposée ici.

La complexité du problème abordé est telle qu'il n'est pas question de proposer des solutions, mais simplement de faire état des difficultés rencontrées, des solutions provisoires retenues, des lacunes qui restent à combler.

## NATURE ET ENVELOPPE DES PROBLEMESPOSES PAR LES GRANDS FICHIERS

Les gros fichiers ont toujours existé dans les entreprises, mais les procédures classiques, manuelles ou mécanographiques, imposaient leur éclatement et leur dispersion géographique.

L'introduction de nouvelles techniques informatiques, les capacités de stockage et les accès qu'elles offrent rend possible aujourd'hui l'étude de fichiers logiques assurant la cohérence des données, l'unicité de la source, une facilité plus grande d'accès à toutes les informations réunies.

Serait prometteuse la voie qu'il, à la fois, rendrait la gestion des données aux utilisateurs concernés et responsables, permettrait aux Directions Générales toutes les synthèses utiles dans des délais acceptables, offrirait au client un service rapide et personnalisé en facilitant le travail fastidieux de l'exploitation du fichier.

Les problèmes sont nombreux à se poser dès que l'on envisage la création d'un tel système :

- qui prend la décision ?
- Quels sont les objectifs de l'entreprise?
- Quels sont les moyens affectés à ce projet ?
- Qui va définir le projet?

Ces options, essentiellement politiques, ne peuvent pas ignorer le cadre technique ; elles doivent notamment répondre à la question: des données pour quoi faire ?

- La banque de données doit-elle être centralisée ou répartie ?
- L'exploitation en temps réel doit-elle se limiter à des consultations ou permettre mise à jour et traitements ?

A partir de là, les inventaires des données nécessaires et des traitements requis sont possibles; c'est une démarche commune en informatique. Mais des problèmes tout aussi ardues vont surgir : ce sont ceux des sauvegardes et surtout des réorganisations des fichiers; les risques d'incidents en cours de réorganisation amenant à mettre en place des procédures spécifiques.

Les problèmes logiciels tournent autour de deux préoccupations essentielles

- Celle d'une structuration des données qui assure leur cohérence, ou leur non redondance mais aussi qui assure une indépendance application donnée et qui permette une évolution raisonnable du contenu,  
celle de la méthode d'accès qui est la relation directe entre les temps d'exploitation et la structuration des données.

c'est là que se pose le problème du choix entre un système de gestion de base des données et le développement d'un logiciel spécifique associé à des méthodes d'accès adaptées.

Cette difficulté de bien définir le cahier des charges de la mise en œuvre d'un gros fichier fait que dans la plupart des cas, on sait maintenant qu'il faudrait réaliser et utiliser un prototype. Ce prototype doit être à la fois modulaire et évolutif, de façon à rapprocher par étapes successives et de façon optimale du service à rendre à l'utilisateur.

Ce qui est acquis aussi, c'est l'importance des moyens matériels nécessaires pour permettre une capacité de stockage suffisante, pour assurer des temps de gestion et d'exploitation raisonnables, tout en offrant les garanties de sécurité requises.

Les matériels mis en œuvre dans les applications qui ont été présentées sont au moins du niveau IBM 370/158 avec des capacités mémoire d'au moins deux mégaoctets. Dans certains cas, même les matériels ont été choisis dans le cadre du projet gros fichier.

Le temps de réponse n'a pas été présenté comme une contrainte lourde, mais les dispositions ont été prises pour que ce problème ne se pose pas, et qu'il ne soit pas nécessaire de quantifier une limite à ne pas dépasser.

En termes d'utilisateur, l'objectif final peut, dans bien des cas, se traduire comme il a été proposé dans un exposé par : "constituer un ensemble de données cohérent où chacun peut puiser à son gré, en conservant une indépendance entre programme utilisateur et organisation des données. Le souci majeur est la moindre maintenance pour permettre une évolution facile et rapide".

Les moyens humains nécessaires sont importants ; en moyens informatiques purs ils dépendent essentiellement du système logiciel retenu.

Ce qui est certain, c'est que l'enjeu économique de telles réalisations est souvent très important et qu'il convient de prendre des garanties, lors de la définition du système et de sa mise en œuvre, mais aussi de son exploitation où il convient de distinguer les notions :

- d'administration du contenant,
- de gestion du contenu,

ce qui a amené des entreprises à définir des équipes d'administration de banques de données.

L'emploi de dictionnaire de données est peu fréquent puisqu'il n'a été cité qu'une fois

Enfin, on ne peut ignorer les problèmes de télétransmission, qu'il s'agisse du cas simple d'une base interrogée à distance, de ceux plus complexes de base mise à jour à distance, et on retrouve là la notion de responsabilité d'administration de donnée. Tout autre est le problème des bases réparties qui a été abordé dans certains exposés.

## ANALYSE DES SYSTÈMES DÉVELOPPÉS DANS LES ENTREPRISES DU CIGREF

### DESCRIPTION SOMMAIRE DES SYSTEMES PRESENTES

Les systèmes développés par les entreprises du CIGREF ayant fait l'objet des exposés dans le cadre de ce groupe de travail, ont leurs caractéristiques essentielles regroupées dans le tableau en annexe du présent rapport.

Il nous a paru intéressant de montrer rapidement quels en étaient les objectifs majeurs :

#### - Le réseau des terminaux agences (R.T.A.) de la Société Générale

Le système ancien était fondé sur la position centralisée des comptes clients et s'appuyait sur un fichier clients et des fichiers par nature de produits (comptes à vue, épargne, titres,...).

L'expérience acquise a permis le développement d'un système nouveau s'appuyant sur une seule base de données interrogeable.

#### - La chaîne des approvisionnements des Houillères du Nord Pas-de-Calais (HBNPC)

Ce système est basé sur deux gros fichiers permanents, celui des stocks, celui des commandes, entrées et factures.

#### - La gestion des logements des HBNPC

Cette application utilise plusieurs fichiers de personnel et un fichier des logements. Son intérêt majeur au sens de ce rapport, réside dans la procédure de mise à jour en temps réel apparent par utilisation d'un fichier intermédiaire de sauvegarde.

#### - Le fichier signalétique des clients du Crédit Lyonnais

Le système en cours de remplacement a pour défaut la duplication des informations signalétiques d'un client ayant des comptes multiples.

Le nouveau système permettra essentiellement de rendre les mises à jour cohérentes à travers la notion de compte-maitre par client, ce qui implique de passer de la notion de compte/client à celle de compte/agence.

#### Le systèmes de réservation électronique des places à la SNCF

Le système actuel prend en compte trois ensemble principaux : le plan de transport (offre de places), les données commerciales affectées à ces offres, les établissements habilités à la commercialisation.

- La gestion des commandes clients à la Redoute

Le système doit permettre de répondre aux commandes de la clientèle dans un premier temps, et d'en assurer le suivi dans un second temps.

La masse des clients (huit millions), l'importance du nombre de mises à jour quotidiennes : 300.000, ont nécessité le développement de logiciels maison performants.

- Le fichier central des impayés à la Banque de France

Ce système a été mis en place pour répondre à la mise en application de la législation en matière de chèques impayés.

Le système permet

- . de répondre aux interrogations des banques,
- de saisir et vérifier les déclarations de chèques impayés en temps réel,
- . de saisir et vérifier les interdictions judiciaires dans les mêmes conditions.

- Le quittancement élaboré à EDF / GDF

Le système mis en place permet la gestion des comptes des clients, la facturation et le recouvrement pour les vingt deux millions d'abonnés à EDF/GDF.

Les abonnés sont répartis en quatre-vingt groupes de facturation chacun d'eux correspondant aux abonnés relevés le même jour. Quatre centres de traitement assurent le traitement de quatre régions.

CARACTERISTIQUES DE L'ORGANISATION DES DONNEES

Dans ce chapitre, nous examinerons les données sous le double aspect de la taille des fichiers qui les renferment (effet de masse) et de leur structure logique (facteur de complexité).

Le problème posé était en effet celui de la maîtrise conjointe

- du volume des données, en temps de traitement mais aussi sécurité, confidentialité.
- de la structure naturellement complexe dans un contexte multiutilisateurs.

Le volume et l'organisation des données

Une classification des fichiers supportant les applications décrites fait apparaître trois groupes distincts :

- celui des 200 à 400 mégaoctets (4 fichiers).
- celui des J.000 à 2.400 mégaoctets (4 fichiers).
- le gros fichier EDF : 25 milliards d'octets (répartis).

Rappelons qu'il ne s'agit pas là de fichiers monolithiques comme ceux mis en œuvre dans les applications de gestion courante.

Bien sûr, ces gros fichiers ont amené toutes les sociétés à maîtriser le problème des multi-volumes. Mais pour réduire le volume des données à manipuler, une seule d'entre elles s'est engagée dans les techniques de compression, de tassement des enregistrements sur disque, cryptage de caractère sur moins de huit bits (compromis temps CPU - espace occupé).

Pour maîtriser l'effet de masse, les entreprises ont surtout cherché des solutions de découpage permettant néanmoins de sauvegarder l'homogénéité du contenu •

### . *Découpage logique*

Une façon de résoudre le problème est de procéder à un découpage logique, les parties correspondant à des fonctionnalités différentes (fichiers commandes et stocks ; fichiers adresses, commandes, paiements).

Cette solution s'inspire des techniques bien connues des fichiers d'applications, avec les défauts de cloisonnements et de dépendance données-traitements.

### *Découpage géographique*

C'est une solution adoptée par des entreprises à couverture nationale avec des agences jouissant d'une certaine autonomie de fonctionnement (Banque) ou d'EdF-GdF mettant en place une véritable distribution au niveau même des quatre vingt dix centres de gestion, ce qui permet de personnaliser les données.

De tels découpages ne peuvent s'envisager dans des cas tels que la réservation des places à la SNCF ou le fichier des impayés de la Banque de France.

Ces fichiers répartis peuvent poser des problèmes délicats lorsqu'une consolidation doit intervenir.

### . *Découpage fonctionnel*

Lorsque le fonctionnement du système d'information est basé sur une notion de cycle (tournées EdF) ou lorsque certaines données sont peu mouvementées (commandes, historiques, ...), lorsque certains traitements peuvent avoir lieu en léger différé (Banque de France), un compromis espace-temps permet souvent d'avoir un système plus performant et plus facilement maîtrisable.

Les avantages de ces découpages sont multiples, ils se prêtent à une mise en place progressive, permettent un fonctionnement en dégradé, une maintenance ponctuelle et des reprises moins lourdes, ils permettent aussi une personnalisation des fichiers, voire à la limite une personnalisation de certains traitements.

Dans le cas de découpages logiques, les parties obtenues ont des structures différentes, nécessairement plus simples que la structure qu'il aurait fallu mettre en œuvre pour obtenir une base unique ; de plus ces structures s'adaptent plus aisément aux impératifs des traitements ultérieurs, surtout s'ils sont bien définis au départ.

La contrepartie de ces découpages pose, dans certains cas, des problèmes sans solution satisfaisante. Il s'agit essentiellement de la difficulté à obtenir des consolidations des informations contenues dans des ensembles différents, et surtout les dangers d'une incohérence entre les données, phénomène accentué encore s'il y a duplication d'informations pour des raisons de performance.

Il faut noter aussi que l'éclatement en fichiers distincts adaptés à des traitements spécifiques pose le problème des interrelations entre ces fichiers avec des difficultés en cas de modification de traitements.

Il apparaît donc qu'avant d'opter pour une solution de découpage, les avantages et inconvénients doivent en être mûrement estimés, des problèmes d'organisation pouvant se substituer aux difficultés d'exploitation dans l'état actuel de la technique informatique.

#### - Structuration et choix d'un SGBD

Dans la plupart des cas, les grands fichiers sont interrogés en temps réel, ils sont multi-applications, et souvent on recherche une indépendance traitements-données.

Ces objectifs sont à la base des systèmes de gestion de bases de données, progiciels commercialisés par les sociétés de services ou les constructeurs.

Toutes les sociétés se sont posé le problème du choix d'un tel progiciel, la démarche commune ayant été l'examen du progiciel constructeur et la comparaison avec un produit portable d'une autre source.

Lorsqu'un SGBD a été mis en œuvre, c'est toujours celui du constructeur, les autres systèmes sur le marché ayant le désavantage de ne pas être suffisamment répandus ou d'être trop frustrés.

A travers ce choix de SGBD, les responsables du traitement des données espéraient résoudre les problèmes informatiques découlant des objectifs fixés, et notamment ceux de la structuration logique des données, permettant interrogation et développement d'applications nouvelles, facilitant la maintenance tout en garantissant la cohérence des données.

A l'usage, il est apparu que les progiciels IDS et IMS, les seuls effectivement utilisés, étaient fiables, et sûrs, qu'ils étaient évolutifs. Les temps de réponse en temps réel sont bons et les pointes convenablement absorbées. Mais les inconvénients sont nombreux et souvent critiques ils sont gourmands en ressources et incitent à des augmentations de configuration, ils sont lourds (souvent trop ambitieux) et peu performants (surtout sur les gros volumes de données), souvent mal documentés.

Comme les SGBD sont intégrés dans la bibliothèque du constructeur, ils en utilisent les méthodes d'accès qui sont souvent inadaptées aux problèmes posés par les gros fichiers et qui trouvent une solution à travers une restructuration des données tenant largement compte des traitements demandés, ce qui est en partie contradictoire avec leur objectif fondamental.

Pour toutes ces raisons, et notamment chaque fois que l'on se heurtait à des problèmes de temps de réponse, ou de trop de généralité du progiciel comparativement à la simplicité des traitements envisagés, la solution adoptée a été celle de développer des produits *maison*.

si le volume des données est un révélateur de certaines insuffisances des outils informatiques disponibles à ce jour, d'autres difficultés résultent de la complexité de la structuration des données. Si les structures arborescentes, les structures en anneau existent dans les progiciels, il apparaît qu'on doit se restreindre à une structuration simple des données : il est remarquable de constater que, dans les applications décrites, les concepteurs se sont limités à des structures à trois niveaux, que le logiciel soit *maison* ou progiciel.

## VIE OPÉRATIONNELLE DU GRAND FICHER

Les choix fondamentaux étant faits (SGBD ou produit maison, répartition géographique ou logique des données, priorité temps réel/traitements différés, aspects d'indépendance des données par rapport aux traitements en fonction de la connaissance que l'on en a...), le fichier est créé, il est utilisé.

Il faut maintenant le faire vivre, garantir la pérennité et l'intégrité des données dans un souci de satisfaction des objectifs du système dans les meilleures conditions économiques possibles.

La terminologie informatique des SGBD retient deux fonctions : celle de la gestion du contenu ou encore gestion des données, et celle de gestion du contenant appelée administration de la base ou du fichier.

### LA GESTION DES DONNEES

C'est là essentiellement un problème utilisateur : les données vivent, il faut saisir les données nouvelles, les intégrer dans le fichier, en sortir les informations obsolètes, valider les opérations correspondantes puis mettre les données contrôlées au service de l'utilisateur final soit pour des interrogations, soit pour des traitements plus élaborés.

La saisie des données se fait essentiellement de deux manières différentes : en traitements par lots, donc dans le cadre d'une application spécifique qui comporte un circuit de validation et de correction, ou en temps réel, donc élément par élément avec possibilité d'un contrôle immédiat pour interrogation sur la transaction qui vient d'être opérée.

La mise à jour : les données nouvelles étant saisies et validées, il faut les intégrer dans le fichier.

A une saisie de type traitement par lots, correspond presque toujours une mise à jour du fichier en traitement par lots donc en différé. Dans les applications présentées, ces mises à jour sont quotidiennes, elles se font de manière séquentielle. Ce procédé est sûr, mais on ne travaille pas sur des données fraîches.

La mise à jour est plus délicate lorsqu'elle se fait en temps réel (risque de détériorer une information). De plus, les reprises en cas d'incident matériel ou logiciel en cours de mise à jour posent quelques problèmes. Des procédures particulières ont été mises en place, qui consistent, en complément

de copies systématiques, à conserver l' historique des transactions, ou à effectuer les mises à jour sur des fichiers considérés comme extensions temporaires du fichier de base, l'interaction définitive n'intervenant que postérieurement.

L'exploitation du fichier, que ce soit en interrogation en temps réel ou en traitement par lots dans le cadre de chaînes d'application ne pose pas de problème bien spécifique en dehors de durées d'exploitation dépendant à la fois de la taille des fichiers et de la complexité de la structuration des données traitées.

- Le dictionnaire des données a d'abord un rôle d'inventaire: il rassemble les noms et les définitions de toutes les données contenues dans le fichier. C'est une base de dialogue entre utilisateurs d'abord, entre utilisateurs et informaticiens ensuite, avant d'être un outil informatique. Les orateurs, dans le cadre de cette étude, s'accordent pour dire qu'un dictionnaire de données est indispensable à la fois à l'utilisateur, au gestionnaire des données et à l'administrateur de la base. Cependant, son concept encore nouveau, fait que ce type d'outil n'a été que peu formalisé dans les entreprises ayant exposé leur problème dans le cadre de ce rapport.

De fait, toutes les entreprises éprouvent les mêmes difficultés pour rendre les utilisateurs réellement responsables de leurs données, ce qui n'est souvent qu'une séquelle fâcheuse du traitement par lots classique.

## L'ADMINISTRATION DE LA BASE

Cette fonction, dont la nécessité est apparue avec la notion de base commune de données, recouvre essentiellement des travaux de nature informatique, permettant de conserver l'intégrité des fichiers en les protégeant contre les malveillances éventuelles mais surtout contre les erreurs d'exploitation ou de manipulation. Il faut aussi, périodiquement, réorganiser les fichiers.

- Les sauvegardes restaurations sont généralement très lourdes, par exemple IMS demande environ une demi-heure pour sauvegarder une base de 200 mégaoctets. De son côté, La Redoute estime à la % la part des sauvegardes dans le coût d'exploitation. Il y a là une étude à conduire pour évaluer le montant de l'assurance qu'il faut prendre pour assurer un seuil de risque. On peut notamment utiliser des procédures de sauvetage rapide sur bande, mais il faudra restructurer la base en cas d'incident.

D'autres aspects de la sécurité consistent:

- à mettre en place des procédures de points de contrôle et de reprise,
- à procéder à des traitements morcelés,

à ,conserver un historique permettant, sur un certain nombre de générations, une reconstitution du fichier.

- à effectuer des mises a Jour en double simultanées,

\_ La réorganisation consiste à restructurer un fichier ou une base que des mises à jour successives, ont désorganisé physiquement, ce qui amène à manipuler un espace disque plus important que nécessaire, tout en compliquant à l'excès la gestion des débordements. Les réorganisations sont lourdes et prennent beaucoup de temps, ce qui conduit souvent à les planifier durant les week-ends, ce qui pose des problèmes au personnel.

\_ La confidentialité : il s'agit de protéger les données ayant un caractère confidentiel, contre des accès non autorisés. La protection actuelle consiste à utiliser des mots de passe connus du système et/ou l'identification du terminal associé. Les habilitations peuvent concerner soit des parties du fichier, soit des sous-structures de la base.

\_ La sécurité est surtout liée à la possibilité de mettre à jour (donc d'effacer ou de polluer) sans contrôle (ou sans procédure en deux temps). Le risque de dégradation de la base est accru si les utilisateurs disposent d'un langage de requête ouvert. Un remède consiste à ne mettre à la disposition des utilisateurs qu'un catalogue de macro-instructions adaptées à leurs besoins, les langages de requête étant strictement réservés aux administrateurs et gestionnaires de la base.

o

## CONCLUSION

De cet examen des problèmes rencontrés lors de l'exploitation de gros fichiers, résulte l'obligation de se poser la question de l'opportunité de la création d'une fonction de gestion des données et d'administration de la base.

L'opinion des ingénieurs de projets qui ont été entendus est qu'il s'agit d'une nécessité dans le cas d'une multiplicité d'utilisateurs et d'applications. Il faut dans ce cas que les responsabilités d'administration et de gestion soient clairement attribuées et centralisées, alors que ces fonctions sont actuellement dispersées entre l'exploitation, les équipes système, les chefs de projets d'application et les divers utilisateurs. Cette dispersion pose des problèmes délicats pour, en cas d'incidents, réunir les compétences et partager les responsabilités après analyse objective des faits.

## CONCLUSIONS GÉNÉRALES

Les réalisations exposées par les participants au groupe de travail conduisent à constater que toutes s'appuient sur des ensembles de données pour lesquelles les difficultés de mise en œuvre et d'exploitation résultent plus de la complexité de structuration de ces données que de leur volume qui n'intervient, de manière très significative il est vrai, que comme amplificateur des difficultés d'exploitation et de maintenance.

La nécessité d'une structuration des données résulte essentiellement des exigences d'interrogation en temps réel (accès - temps de réponse), et d'exploitation multi-applications (indépendance traitements - données). Un autre objectif est de garantir une maintenance facile permettant de répondre aux évolutions à structure de données figées.

Les responsables de projets se sont donc tout naturellement tournés vers les Systèmes de Gestion de Bases de Données, et les choix montrent un souci d'aller vers l'outil le plus simple qui puisse répondre au problème posé; les systèmes à vocation générale proposés par les sociétés de service ont été écartés

au profit des systèmes "constructeurs" et plusieurs réalisations ont été faites sur des produits développés sur mesure dans les entreprises pour s'affranchir de toute la lourdeur liée aux spécifications générales jugées inutiles sur ces progiciels.

c'est dire l'importance que revêt la phase de conception d'un tel système: ce n'est pas un problème d'informaticien mais un problème d'entreprise, et on peut regretter que l'aspect méthodologie de la définition des spécifications du projet n'ait jamais été abordé au cours des exposés.

Si les besoins des utilisateurs semblent avoir été satisfaits dans une large mesure, du moins pour ceux initialement exprimés, à l'usage apparaissent des modifications de la demande qui ne peuvent être prises en compte facilement que s'il n'y a pas remise en cause de la structure des données choisies. Les Systèmes de Gestion de Base de Données apportent en effet une inertie importante à des évolutions dans la structure des données qui reste, quoiqu'on en dise, fortement dépendante de certains traitements.

Toutes ces considérations expliquent que les responsables de ces grands projets considèrent que ceux-ci doivent faire l'objet d'une première version prototype dont on observe le fonctionnement opérationnel suffisamment longtemps pour permettre une critique constructive des utilisateurs et qu'on en tire les enseignements au niveau des contraintes d'exploitation.

L'examen des conditions de vie de ces grands fichiers a conduit à augmenter la responsabilité de l'utilisateur final dans son domaine de compétence: la qualité des données et leur sécurité. C'est dans ce sens que va la distinction qu'il faut faire entre une fonction informatique :

- la gestion du contenant (intégrité physique des données, sécurité, contrôle des accès, ...) et une fonction du strict ressort de l'utilisateur:

la gestion du contenu, ou plus simplement gestion qui comporte tous les problèmes de définition des données, de leurs relations, de leur validation, de leur utilisation.

Ces deux notions prennent d'autant plus d'importance que le nombre d'utilisateurs est grand, ce qui implique la désignation d'entités responsables séparément de ces aspects vitaux pour la vie de l'application.

TABLEAU RECAPITULATIF DES EXPOSES FAITS SUR LE THEME : GESTION DE GRANDS FICHIERS

SOCIETE ORATEUR	RÉF C/R	APPLICATION	DESCRIPTION DU FICHER				EXPLOITATION DU FICHER				OBS.
			VOLUME Enreg. Octets en millions	ORGANISAT. Trad. SCBD	Nbre niveau	Décou- page	Mise à jour batch Taux	Interro. Batch T.R.	Hbr. util. 1 N	Cellu- le adm.	
SOCIETE GENERALE M. VOISIN	2/76	Réseau de terminaux AGENCES	3 1800 2100	IDS IDS	2 2		X 2% 50%	X X X	X X X		Batch T. réel
HOUILLERES NORD P. de CALAIS MM. BORE HOUDIN	3/76	Stocks com- mandes Logements	0,12 200 250	X	r		X X	X X	X X		
CREDIT LYONNAIS M. VITRY	4/76	Fichier si- gnalétique clients	2,7 1000	BDAM	2	3 sites à 3 lots	X	X	X		-
S.N.C.F. M. RIFFAULT	5/76	Syst. de réservation électronique des places	150 1500	IMS	3 2	20 bases	X X	X X	X X		offre réservation
LA REDOUTE Melle CATRICE	6/76	Gestion de commandes clients	8 2400	X		8 bases	X 5%	X	X	X	X
BANQUE DE FRANCE MM. LESPAGNOL JOIE	1/77	Fichier cen- tral chèques impayés	0,7 360	IMS	2		10% X	X	X	X	
E.D.F. M. PARR	2/77	Quittancement élaboré	22 25000	X	3	4 sites x 80 lots 90 sites x 80 lots	X	X	X		fichier bande